BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Molecular Cloning

A LABORATORY MANUAL

T. Maniatis Harvard University
E. F. Fritsch Michigan State University
J. Sambrook Cold Spring Harbor Laboratory

Cold Spring Harbor Laboratory
1982

Molecular Cloning

A LABORATORY MANUAL

Printed in the United States of America
Book and cover design by Emily Harste

Front cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. *Thomas R. Broker, Louise T. Chow, and James I. Garrels*

Back cover: E. coli DH1 with fimbriae was negatively stained with phosphotungstic acid and the electron micrograph was digitized and assigned false color by computer. *Jeffrey A. Engler, Thomas R. Broker, and James I. Garrels*

Cataloging in Publications data

Maniatis, T.
Molecular cloning.

(A laboratory manual)
Bibliography: p.
Includes index.
1. Molecular cloning. 2. Eukaryotic cells. I. Fritsch, Edward F. II. Sambrook, Joseph. III. Title. IV. Series.
QH442.2.F74 574.87'3224 81-68891
ISBN 0-87969-136-0 AACR2

12 Vectors That Express Cloned DNA in *Escherichia coli*

Expression vectors contain sequences of DNA that are required for the transcription of cloned copies of genes and the translation of their mRNAs in *Escherichia coli*. Such vectors have been used both to express eukaryotic genes in *E. coli* and to increase production of prokaryotic gene products.

The three major requirements for expression of a cloned gene in *E. coli* are:

1. The coding region of the gene must not be interrupted by intervening sequences.

2. The gene must be placed under the control of an *E. coli* promoter that is efficiently recognized by *E. coli* RNA polymerase.

3. The mRNA must be relatively stable and efficiently translated. In addition, to be recovered, the foreign protein produced in *E. coli* must not be rapidly degraded by bacterial proteases.

PROMOTERS

A promoter is a DNA sequence that directs RNA polymerase to bind to DNA and to initiate RNA synthesis. Different promoters work with different efficiencies: Strong promoters cause mRNAs to be initiated at high frequency; weak promoters direct the synthesis of rarer transcripts. Comparison of the sequences of a number of different promoters reveals two highly conserved regions, one located about 10 bp (−10 region or Pribnow box; Pribnow 1975) and the other about 35 bp (−35 region) upstream from the point at which transcription starts (for a review, see Rosenberg and Court 1979). These two regions are thought to be important in determining promoter strength because mutations that decrease the frequency of transcription usually decrease the amount of homology with the conserved sequences. However, other, more moderately conserved regions of promoters may also contribute to promoter strength. Furthermore, the number of nucleotides that separate the conserved sequences is important for efficient promoter function. For example, 16 to 19 nucleotides separate the −10 region from the −35 region; mutations altering the spacing between these two conserved regions in a *lac* promoter (*lacP*^{s}; de Crombrugghe et al. 1971; Stefano and Gralla 1982) and in the β-lactamase promoter (Jaurin et al. 1981) change the "strength" of the promoter. These results indicate either that there is an optimal spacing common to all promoters or that there is an optimal spacing for any individual promoter that is dependent on its particular DNA sequence.

The only true test of the efficiency of a promoter is to measure the frequency with which the synthesis of the appropriate mRNA is initiated. Because this value is difficult to obtain from in vivo studies, the efficiency of a promoter is frequently deduced indirectly from the level at which the relevant protein product is expressed. However, a major difficulty in comparing the strengths of promoters by this method is that different mRNAs contain different untranslated leader sequences at their 5′ termini that may affect the efficiency of translation of the mRNA. The level of expression of the protein product may therefore be due to the strength of the promoter, to the composition and length of the untranslated leader sequence, or to a combination of the two.

Despite these uncertainties, it is well-established that many *E. coli* genes are controlled by relatively weak promoters. Thus the expression of such genes can be increased by placing them downstream from an efficient promoter (e.g., *lac*uv5, *trp*, *tac*, *trp-lac*uv5 hybrid promoter, λp_L, *ompF*, *bla*). Eukaryotic promoters function extremely poorly, if at all, in *E. coli*, and efficient expression of eukaryotic proteins has been achieved only when the coding sequence is placed under the control of a strong *E. coli* promoter.

The most useful promoters for expressing foreign genes in *E. coli* are those that are both strong and also regulated. Obviously if the product of the cloned gene is toxic to *E. coli* cells, then coupling the gene to a strong, unregulated promoter is not desirable. In addition, a constitutively high level transcription may interfere with plasmid DNA replication and lead to plasmid instability (Remaut et al. 1981). The presence of efficient termina-

tors of transcription placed downstream from the promoter may circumvent this problem. In fact, strong promoters (e.g., certain bacteriophage T5 promoters) require the presence of a strong downsteam termination signal if they are to be stably maintained on a plasmid (Gentz et al. 1981).

Vectors Using the p_L Promoter of Bacteriophage λ

The p_L promoter of bacteriophage λ is a strong, well-regulated promoter that has been used in several expression vectors (Hedgpeth et al. 1978; Bernard et al. 1979; Remaut et al. 1981; Shimatake and Rosenberg 1981). A gene encoding a temperature-sensitive λ repressor (e.g., λ*cIts*857) may either be included in the cloning vector or may be provided by a prophage resident in the bacterial chromosome (Bernard et al. 1979). At low temperature (31°C), the p_L promoter is maintained in a repressed state by the *c*I-gene product. After the activity of the repressor is destroyed by raising the temperature of the culture, the p_L promoter directs the synthesis of large quantities of mRNA. Several vectors utilizing the λp_L promoter are described below.

pPLa2311

Insertion of DNA into the *Pst*I site of pPLa2311 (Remaut et al. 1981; see Fig. 12.1) inactivates the plasmid's ampicillin-resistance (*bla*) gene. Transformed cells that are resistant to kanamycin and sensitive to ampicillin are therefore likely to have insertions of foreign DNA at the *Pst*I site. The λp_L promoter directs the synthesis of large amounts of mRNA encoding the gene product to be expressed when the foreign DNA is inserted in the correct orientation.

pPLa8

pPLa8 was constructed by inserting a *Bam*HI linker in the *Pst*I site of pPLa2311 (Remaut et al. 1981). Bacteria carrying pPLa8 are therefore sensitive to ampicillin, so that insertional inactivation cannot be used to screen for recombinant plasmids that carry foreign DNA at the *Bam*HI site.

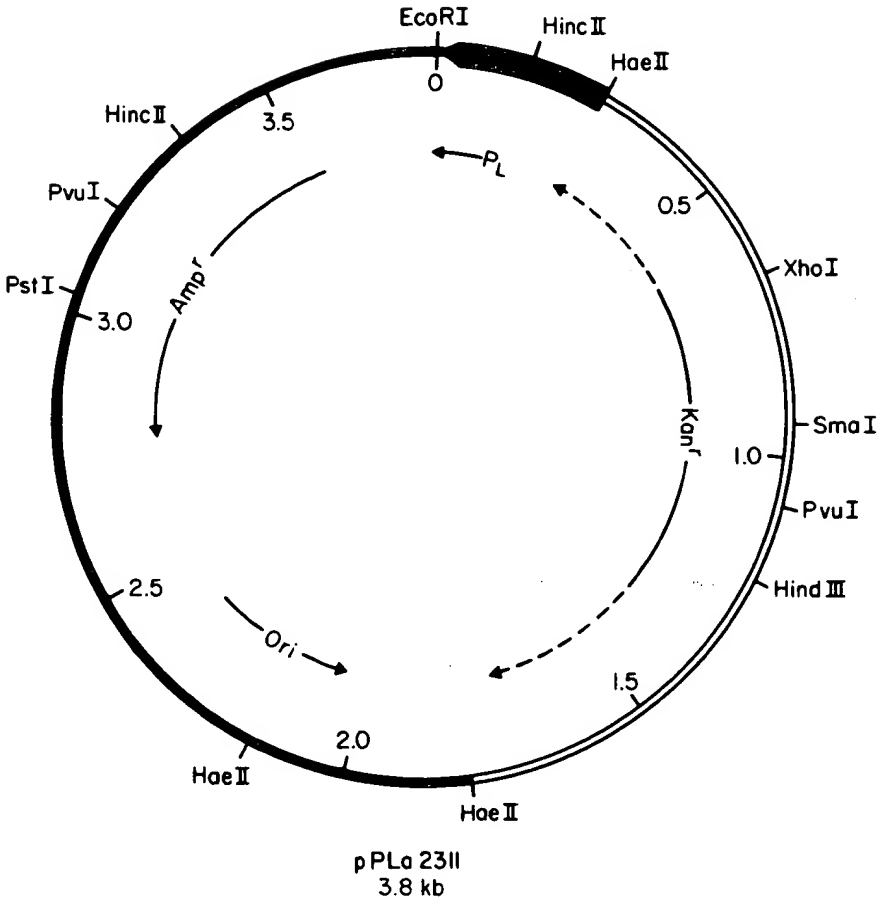


Figure 12.1

pPLa2311, a plasmid 3.8 kb in length with unique *Eco*RI and *Pst*I sites, that carries the p_L promoter of bacteriophage λ. The heavy arrow indicates the location of the p_L region and the direction of transcription. The solid lines represent DNA derived from pBR322. The open line shows the region carrying the gene that confers kanamycin resistance (whose reading direction is unknown). Insertion of foreign DNA into the *Pst*I site inactivates the ampicillin-resistance gene. Genes inserted into either the *Pst*I site or the *Eco*RI site can be regulated by introducing the recombinant plasmid into a temperature-sensitive λ lysogen (*cIts857*). The cells carrying this plasmid are grown to late log phase at 32°C and are then shifted to 42°C to inactivate the *c*I-gene product and to turn on the p_L promoter. Plasmid pPLa8 has the same structure as pPLa2311, except that a *Bam*HI linker has been inserted into the *Pst*I site (Remaut et al. 1981).

pKC30

The plasmid pKC30 (Shimatake and Rosenberg 1981) has been shown to direct the production of large amounts of a toxic protein (in this case, the bacteriophage λ *c*II-gene product; see Fig. 12.2) which under normal circumstances is produced by bacteriophage λ in low amounts and is rapidly degraded when present in small amounts.

The plasmid contains an *Hpa*I site located 321 bp downstream from the bacteriophage λ p_L transcription start site (see Fig. 12.2). When the bacteriophage λ *c*II gene was cloned into this site in the correct orientation, a lysogen (in which the resident prophage was making λ repressor) could be transformed at high efficiency, while a nonlysogen could not be transformed (Shimatake and Rosenberg 1981). In the nonlysogen, there was no *c*I repressor to reduce synthesis of the *c*II-gene product, and the constitutive synthesis of large amounts of the protein proved to be lethal for *E. coli*. The fact that a lysogen was able to tolerate the plasmid indicates that an integrated copy of bacteriophage λ makes enough repressor to reduce the expression of the *c*II gene to a nonlethal level.

Large amounts of the *c*II-gene product (4% of total cell protein) could be produced from this plasmid by raising the temperature of a lysogen in which the defective prophage carried a temperature-sensitive mutation in the repressor gene (λ*cIts*857). The bacteriophage λ *N*-gene product was also required for high levels of production of *c*II protein; presumably the *N* protein blocked termination of transcription at t_{R1}, a rho-dependent transcription termination site located upstream from the *c*II gene (see Chapter 1).

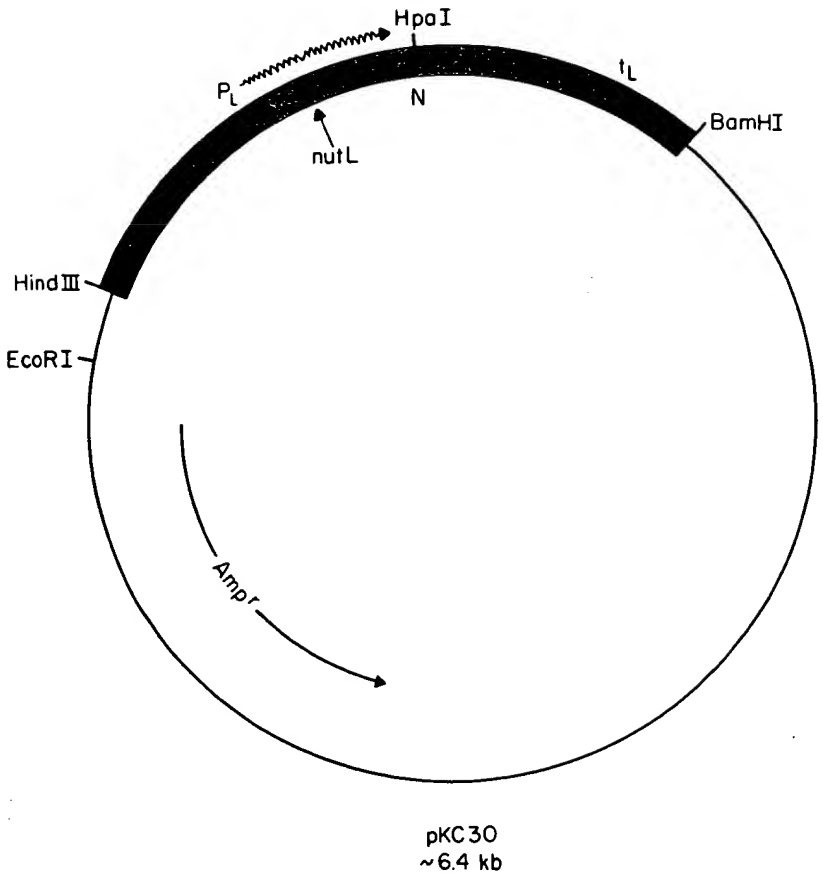


Figure 12.2

pKC30, a plasmid ~ 6.4 kb in length that carries the p_L promoter of bacteriophage λ and an *Hpa*I recognition site located 321 nucleotides downstream from the p_L transcriptional start site. The plasmid is a derivative of pBR322 and contains a *Hin*dIII-*Bam*HI fragment (black box) derived from bacteriophage λ inserted between the *Hin*dIII and *Bam*HI sites within the tetracycline-resistance gene. The insertion contains the promoter signal, p_L, a site recognized by the *N*-gene product (*nut*L), the *N* gene itself, and the strong rho-dependent transcription-termination signal t_L. The *Hpa*I recognition site lies within the coding region of the *N*-gene coding region. Sequences inserted into the *Hpa*I site can be regulated by introducing the recombinant plasmid into a temperature-sensitive λ lysogen (*cIts*857). The cells are grown to late log phase at 32°C and then shifted to 42°C to inactivate the *c*I-gene product and to turn on the p_L promoter. This vector has been used to express the λ*c*II protein at a level such that the protein comprises 4% of the total protein of the cell (Shimatake and Rosenberg 1981).

Other Promoters

Expression of cloned genes in bacteria can also be controlled by other promoters. For example:

1. The *ompR* gene codes for a positive regulatory protein that controls the expression of *ompF*, a gene coding for a major outer membrane protein of *E. coli*. A cold-sensitive mutant mapping in the *ompR* gene has been isolated. Therefore, the transcription from the *ompF* promoter can be activated by raising the temperature of the culture (T. Silhavy, pers. comm.).

2. The *trp* promoter is regulated by *trp* repressor and can be induced by the addition of 3-indolylacetic acid to the medium (Morse et al. 1970) or by tryptophan starvation (see Miller and Reznikoff 1978).

3. The *lac* promoter is regulated by *lac* repressor and can therefore be induced by the addition of the inducer isopropyl-β-D-thiogalactoside (IPTG) to the bacterial culture (see Miller and Reznikoff 1978).

4. Finally, a hybrid promoter consisting of the *trp* −35 region fused to the *lac* −10 region and the *lac* operator has been constructed by de Boer et al. (1982). This strong *trp-lac* hybrid promoter (or *tac* promoter) is regulated by the *lac* repressor.

As discussed above, increased expression of an *E. coli* protein encoded by a gene that does not have a strong promoter may be achieved by placing the gene downstream from one of these or other strong *E. coli* promoters (Selker et al. 1977; Backman and Ptashne 1978).

RIBOSOME-BINDING SITES

To achieve high levels of gene expression in *E. coli*, it is necessary to use not only strong promoters to generate large quantities of mRNA, but also ribosome-binding sites to ensure that the mRNA is efficiently translated. In *E. coli*, the ribosome-binding site includes an initiation codon (AUG) and a sequence 3–9 nucleotides long located 3–11 nucleotides upstream from the initiation codon (Shine and Dalgarno 1975; Steitz 1979). This sequence, which is called the Shine-Dalgarno (SD) sequence, is complementary to the 3′ end of *E. coli* 16S rRNA. Binding of the ribosome to mRNA is thought to be promoted by base pairing between the SD sequence in the mRNA and the sequence at the 3′ end of the 16S rRNA (Steitz 1979).

The efficiency of translation of an mRNA could be affected by several factors:

1. The degree of complementarity between the SD sequence and the 3′ end of the 16S rRNA.

2. The spacing and possibly the DNA sequence lying between the SD sequence and the AUG (Roberts et al. 1979a,b; Guarente et al. 1980a,b). The level of expression of genes has been measured in plasmids in which this was systematically altered; an optimal spacing between the *trpL* SD sequence and the ATG of two genes was determined (D. Goeddel, unpubl.). Comparison of different mRNAs shows that there are statistically preferred sequences from positions −20 to +13 (where the A of the AUG is position 0) (Gold et al. 1981). Leader sequences have been shown to influence translation dramatically (Roberts et al. 1979a,b).

3. The nucleotide following the AUG, which affects ribosome binding (Taniguchi and Weissman 1978).

EXPRESSION OF EUKARYOTIC GENES

Many of the vectors described below are designed so that genes can be placed downstream from a promoter. However, for efficient expression of a eukaryotic gene, a bacterial ribosome-binding site also must be provided. This can be accomplished in the following two ways.

Vectors That Express Unfused Eukaryotic Proteins

The first method of providing a bacterial ribosome-binding site involves the synthesis of a protein that initiates at the AUG of the eukaryotic mRNA and that contains no bacterial sequences at its amino terminus. One way to accomplish this is, first, to introduce a restriction site immediately upstream from the ATG of the gene to be expressed and then to clone the gene directly into a compatible restriction site immediately downstream from an SD sequence (Goeddel et al. 1979a, 1980b; Edman et al. 1981). A "hybrid" ribosome-binding site (Backman and Ptashne 1978) is thus constructed and is composed of a bacterial SD sequence and the ATG of the gene to be expressed. The protein is synthesized, unfused to any other protein. If the DNA sequence to be expressed lacks an ATG codon, then one must be provided by chemical DNA synthesis (e.g., Goeddel et al. 1979a; Davis et al. 1981).

For example, the "mature" form of human growth hormone (HGH) does not begin with a methionine, because normally it is synthesized as a precursor with an aminoterminal signal peptide that is cleaved during secretion. To construct a plasmid expressing the mature form of HGH in *E. coli*, chemically synthesized DNA containing an *Eco*RI site, an ATG codon, and the first 24 codons of mature HGH was inserted, along with cDNA encoding the rest of HGH, into a vector with an *Eco*RI site just downsteam from the *lac* promoter and SD sequence. The SD sequence was separated from the chemically synthesized ATG by 11 bp of DNA generated by ligation of the *Eco*RI ends. This plasmid directed synthesis of HGH under control of the *lac* promoter.

An alternative means of introducing an ATG codon into a cloned gene and of maintaining the bacterial ribosome-binding site is to use the vector pAS1, described below. With this vector, the λp_L promoter, the λcII SD sequence, and ATG codon are fused directly to the aminoterminal coding portion of the eukaryotic gene to be expressed.

It may be critical to optimize the distance between the bacterial SD sequence and the ATG of the eukaryotic gene in order to obtain efficient expression of the gene (for review, see Guarente et al. 1980a). Optimal placement of the promoter fragment may be achieved by the methods described in Roberts et al. (1979a,b). In brief, a unique restriction site is positioned in a plasmid within 100 bp upstream from the initiation codon. The plasmid DNA is then cleaved at that restriction site and resected to various extents with an exonuclease. A DNA fragment carrying the promoter and the SD sequence then is inserted to produce a series of plasmids that contain the SD

sequence at varying distances from the initiation codon. Several DNA fragments, termed portable promoter fragments, have been designed for this purpose, including one containing the *lac*uv5 "portable promoter" and the *lacZ* SD sequence that has been used to express several prokaryotic and eukaryotic genes in *E. coli* (see Fig. 12.3) (Backman and Ptashne 1978; Roberts et al. 1979b; Guarente et al. 1980a,b; Taniguchi et al. 1980b).

A number of other vectors for producing eukaryotic proteins that are not fused to bacterial peptides are discussed below.

ptac12

Plasmid p*tac*12, which contains the hybrid promoter *tac* (see Fig. 12.4), can be used in the same manner as the *lac*uv5 portable promoter (E. Amann and J. Brosius, pers. comm.). However, the *tac* promoter is more efficient than the *lac*uv5 promoter and should be used in a bacterial strain containing a mutation in the *lac*I gene that causes the *lac* repressor to be overproduced (*lac*I^q). Even in such a strain, however, transcription from the *tac* and the ordinary *lac* promoters is not fully repressed when many copies of the plasmid are present (for discussion, see de Boer et al. 1982).

Because expression of some foreign proteins is detrimental to *E. coli*, it may be expedient to maximize expression of a foreign gene using the *lac* portable promoter and then to use the plasmid with the optimal SD-ATG spacing to construct a derivative carrying the stronger *tac* promoter. This is best achieved by using *Hpa*II to isolate a DNA fragment (encoding the *lac* Pribnow box, the *lac* leader, and part of the coding region of the gene) from the plasmid in which the *lac*uv5 promoter has been positioned in front of a eukaryotic gene. This fragment is then inserted into the *Cla*I site of the plasmid pEA300 (see Fig. 12.5) (E. Amann and J. Brosius, pers. comm.). The *lac* promoter is thus replaced by a *tac* promoter without altering the leader or the spacing between the SD sequence and the ATG. Plasmid pEA300 has the transcriptional termination signals from the *rrnB* operon located downstream from the site of insertion of the gene.

p*trpL*1

Plasmid p*trpL*1 carries another portable promoter fragment that can be used in the same way as the *lac* and *tac* portable promoters (see Fig. 12.6) (Edman et al. 1981).

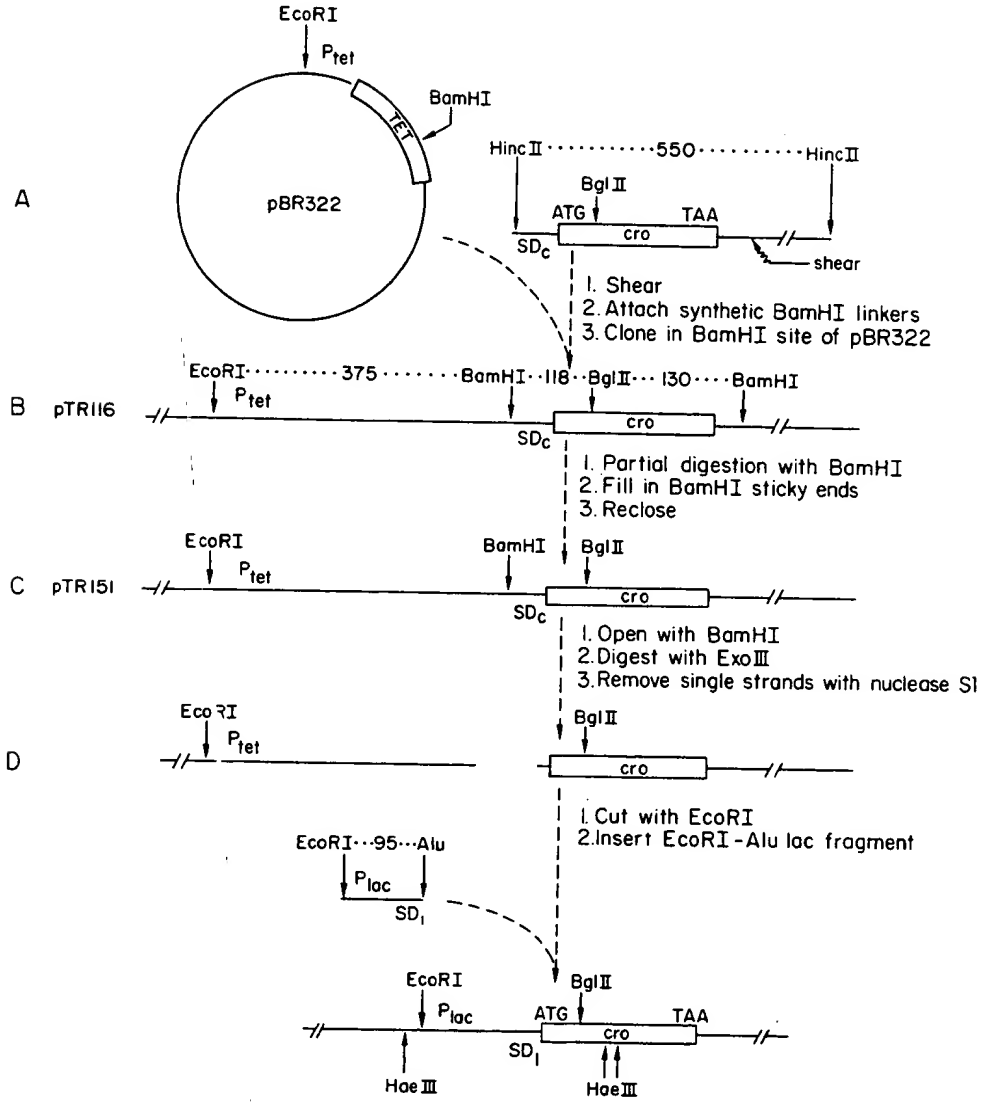


Figure 12.3

Schematic representation of the portable promoter procedure. The approximate locations of several restriction endonuclease cleavage sites are shown for the plasmid pBR322, for a DNA fragment bearing the *cro* gene of phage λ, and for a DNA fragment bearing the promoter of the *lac* operon (see Backman and Ptashne 1978 for the source of this fragment). The locations of the *tet* and *lac* promoters are indicated, as are the extent of the *tet* and *cro* genes. SD_c and SD_l indicate the Shine-Dalgarno sequences of the *cro* and *lacZ* genes, respectively. AUG and UAA are the start and stop signals for translation of the *cro* protein. Distances are indicated in base pairs. *(A)* The DNA fragment bearing the *cro* gene was shortened by shearing to remove certain λ control elements near the 3′ end of the gene, and the smaller fragment was inserted into the *Bam*HI site in pBR322 by using *Bam*HI linkers. *(B)* The *Bam*HI site near the carboxyl terminus of the *cro* gene was eliminated. *(C)* The plasmid was opened at the *Bam*HI site, and varying amounts of DNA were removed by treatment with exonuclease III and nuclease S1. *(D)* The partially resected plasmid was cut at the *Eco*RI site, and the *lac* promoter (bearing the uv5 mutation, rendering it independent of catabolite activator) was inserted by "sticky-end" ligation at its *Eco*RI end and by "blunt-end" ligation to the resected plasmid DNA at its *Alu* end. The efficiency of steps *C* and *D* is fairly high—about 200–400 plasmids result from each microgram of pTR151 used. (Adapted from Roberts et al. 1979a.)

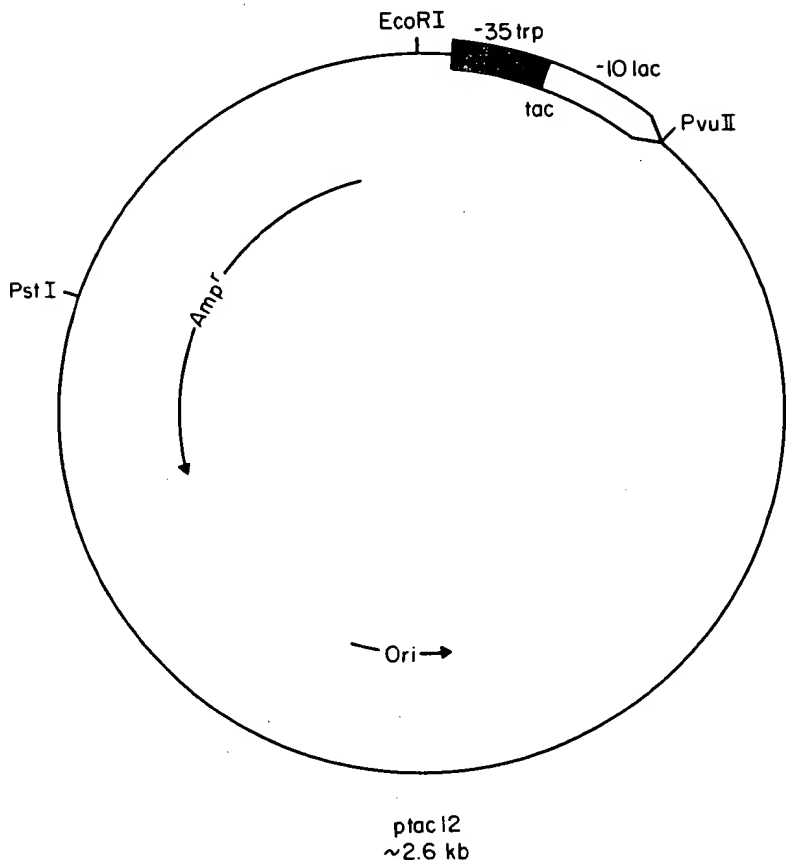


Figure 12.4

p*tac*12, a plasmid ~ 2.6 kb in length that contains the portable hybrid *trp-lac (tac)* promoter (E. Amann and J. Brosius, pers. comm.). A fragment (~ 260 bp) containing the promoter is isolated after digestion of this plasmid with *Eco*RI and *Pvu*II. This DNA encodes the *tac* promoter and the *lac* leader sequences. The *Pvu*II site is located 5 bp downstream from the SD sequence of the *lacZ* gene. The portable promoter fragment is inserted in front of the initiation codon of the gene to be expressed (as outlined in Fig. 12.3). Because the *tac* promoter is strong, recombinant plasmids containing the portable promoter should be carried in a $lacI^q$ strain to repress transcription.

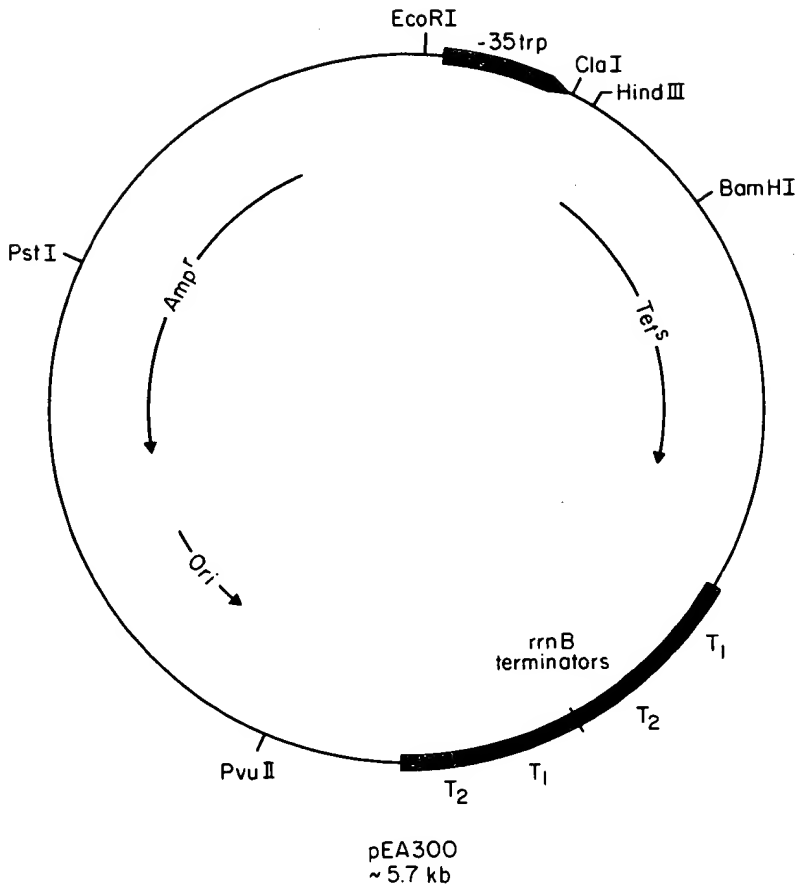


Figure 12.5

pEA300, a plasmid ~ 5.7 kb in length that carries a 192-bp fragment of *trp* DNA sequences cloned into the *Cla*I site of pBR322. By cutting this plasmid with *Cla*I and inserting an *Hpa*II fragment derived from a plasmid carrying the *lac*uv5 promoter, the *tac* promoter is constructed. In addition, pEA300 carries downstream from the promoter two 500-bp fragments, each of which contains two *rrnB* terminators, arranged in tandem (E. Amman and J. Brosius, pers. comm.).

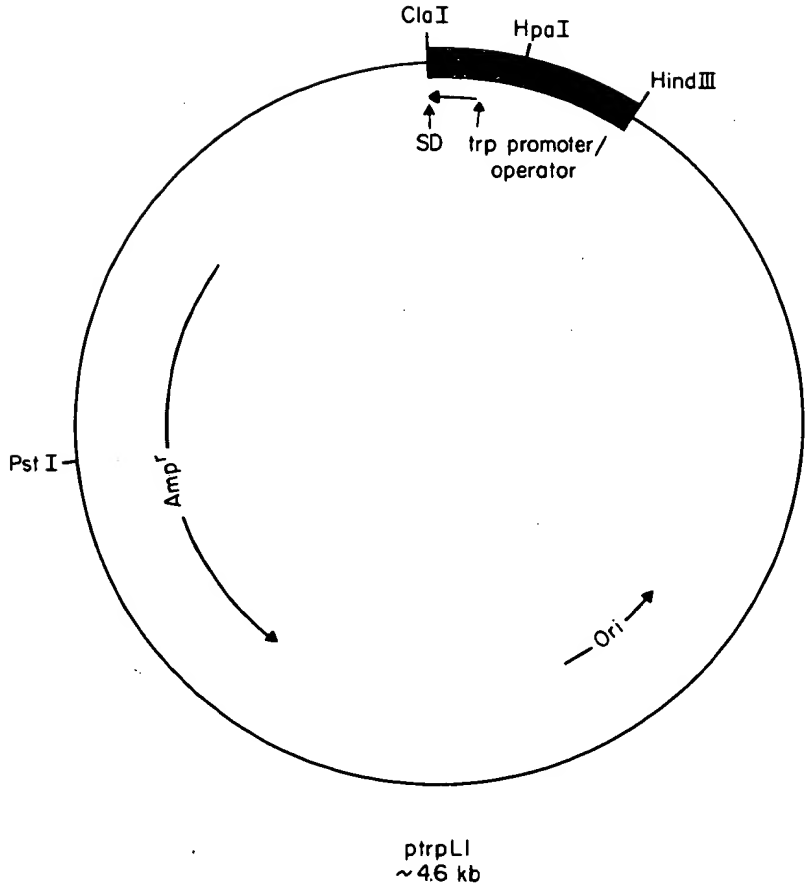


Figure 12.6

p*trpL*1, a plasmid ~ 4.6 kb in length containing a sequence that could be used as a portable *trp* promoter. To use the portable promoter, the plasmid is cleaved at the *Cla*I site (3 bp downstream from the SD sequence of *trpL*). After the protruding termini have been repaired by the Klenow fragment of DNA polymerase, the plasmid is digested with *Hin*dIII and the promoter fragment is isolated. This fragment can be placed in front of the gene to be expressed. Alternatively, the plasmid can be used directly as an expression vector by inserting into the *Cla*I site of the p*trpL*1 plasmid a DNA fragment extending from a restriction site immediately upstream from the ATG of the gene to be expressed through (or into) the coding sequence (Edman et al. 1981).

pAS1

Another system for the expression of unfused foreign genes in *E. coli* utilizes the vector pAS1 (Fig. 12.7; A. Shatzman and M. Rosenberg, pers. comm.), which provides not only a promoter (λp_L) and an SD sequence, but also an ATG codon separated from the SD by its normal distance in the λcII gene. This vector is therefore particularly useful for expressing eukaryotic coding sequences that lack an ATG. pAS1 is cleaved at the *Bam*HI site and digested with mung-bean nuclease to create a blunt end immediately downstream from the ATG. This ATG can then be joined by blunt-end ligation to a DNA fragment coding for the segment of protein to be expressed.

This method requires that a blunt-end be created immediately before a particular codon (Panayotatos and Truong 1981; A. Shatzman and M. Rosenberg, pers. comm.). This can be done as follows (see Fig. 12.8).

1. Insert two unique restriction sites upstream from the coding sequence to be expressed.

   Site 1 must generate after cleavage a protruding terminus, 4 nucleotides long; the sixth nucleotide of this site will eventually become the first nucleotide of the second codon of the protein that will be expressed.

   Site 2 must lie between site 1 and the start of the gene and must be closer to the start of the gene than to site 1.

2. Cleave the DNA at site 2 and resect with nuclease *Bal*31 to produce a population of molecules that terminate close to the desired codon.

3. Cleave the DNA at site 1.

4. Repair the DNA by the Klenow fragment of *E. coli* DNA polymerase I to create a blunt end that contains 5 of the 6 nucleotides of restriction site 1.

5. Recircularize the DNA by ligation and use the plasmid to transform bacteria to antibiotic resistance.

6. Screen individual colonies for the presence of plasmids in which restriction site 1 has been regenerated. Such regeneration occurs when, by chance, digestion with *Bal*31 generates a DNA molecule whose terminus carries the particular base pair required to complete site 1. Included in this subpopulation of plasmids will be those in which the last nucleotide of site 1 is the first nucleotide of the second codon of the gene to be expressed.

7. Cleave one of these plasmids at site 1 and digest with mung-bean nuclease to generate a blunt end directly preceding the first nucleotide of the second codon, and then cleave with a restriction enzyme that cuts downstream from the gene to obtain a DNA fragment that can be inserted after the ATG of plasmid pAS1.

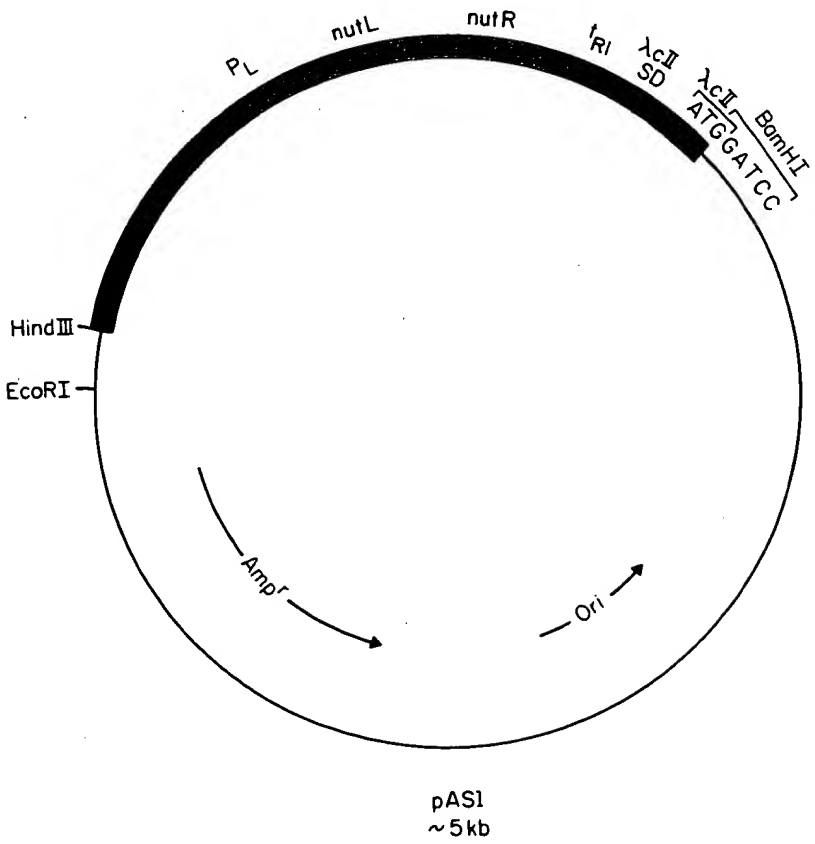
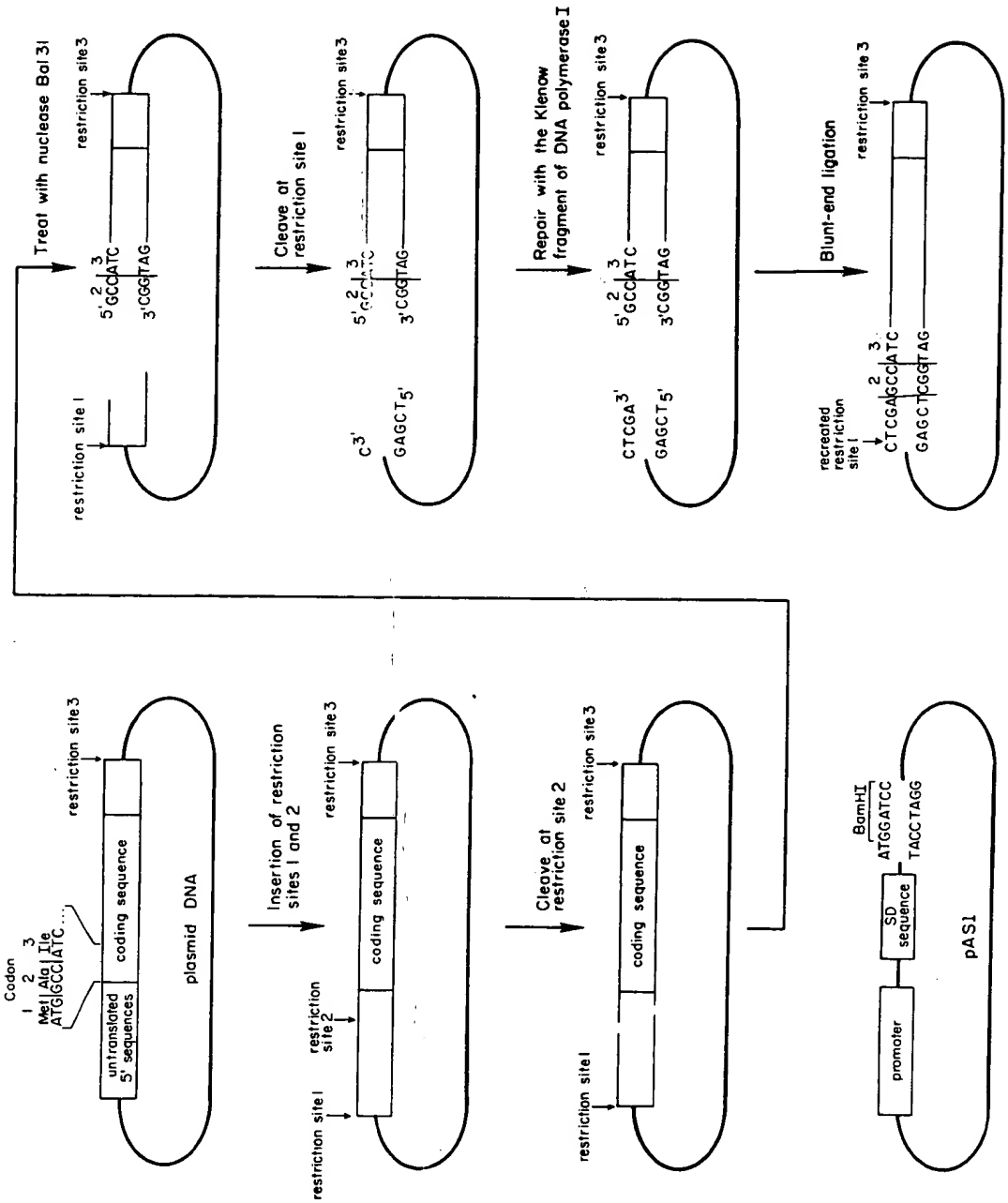


Figure 12.7

pAS1, a plasmid ~ 5 kb in length that carries the bacteriophage λ p_L promoter and a unique *Bam*HI site located at the ATG of the λ*c*II gene. This plasmid is a derivative of pKC30 (Fig. 12.2) into which the λ*c*II gene was inserted at the *Hpa*I site. The *c*II gene was then resected by exonuclease digestion until only the initiation codon ATG remained (the G of the ATG is the first nucleotide of a *Bam*HI site). To express a gene lacking an initiation codon, pAS1 is digested with *Bam*HI and then treated with mung-bean nuclease to remove the protruding, single-stranded termini. Ligation of this blunt-ended DNA to a blunt-ended DNA fragment that begins with the second codon of the gene to be expressed places that gene in-frame with the ATG. Genes inserted in this manner are regulated by introducing the recombinant plasmid into a temperature-sensitive, bacteriophage λ lysogen (*cIts*857). The cells are grown to late log phase at 32°C then shifted to 42°C to inactivate the repressor and to turn on the p_L promoter. The inserted gene can also be regulated by the action of the *N* protein at *nut*L and *nut*R to antiterminate at t_{R1} (A. Shatzman and M. Rosenberg, pers. comm.).


Treat with nuclease Bal 31
restriction site3
restriction site I
5'GCCATC
3'CGGTAG
Cleave at restriction site I
C3'
GAGCT5'
Repair with the Klenow fragment of DNA polymerase I
CTCGA3'
Blunt-end ligation
recreated restriction site I
CTCGAGCCCATC
GAGCTCGGGTAG
Codon 1 2 3
Met Ala Ile
ATGGCCATC...
untranslated 5' sequences
coding sequence
plasmid DNA
Insertion of restriction sites I and 2
restriction site2
Cleave at restriction site 2
BamHI
ATGGATCC
TACCTAGG
promoter
SD sequence
pAS1

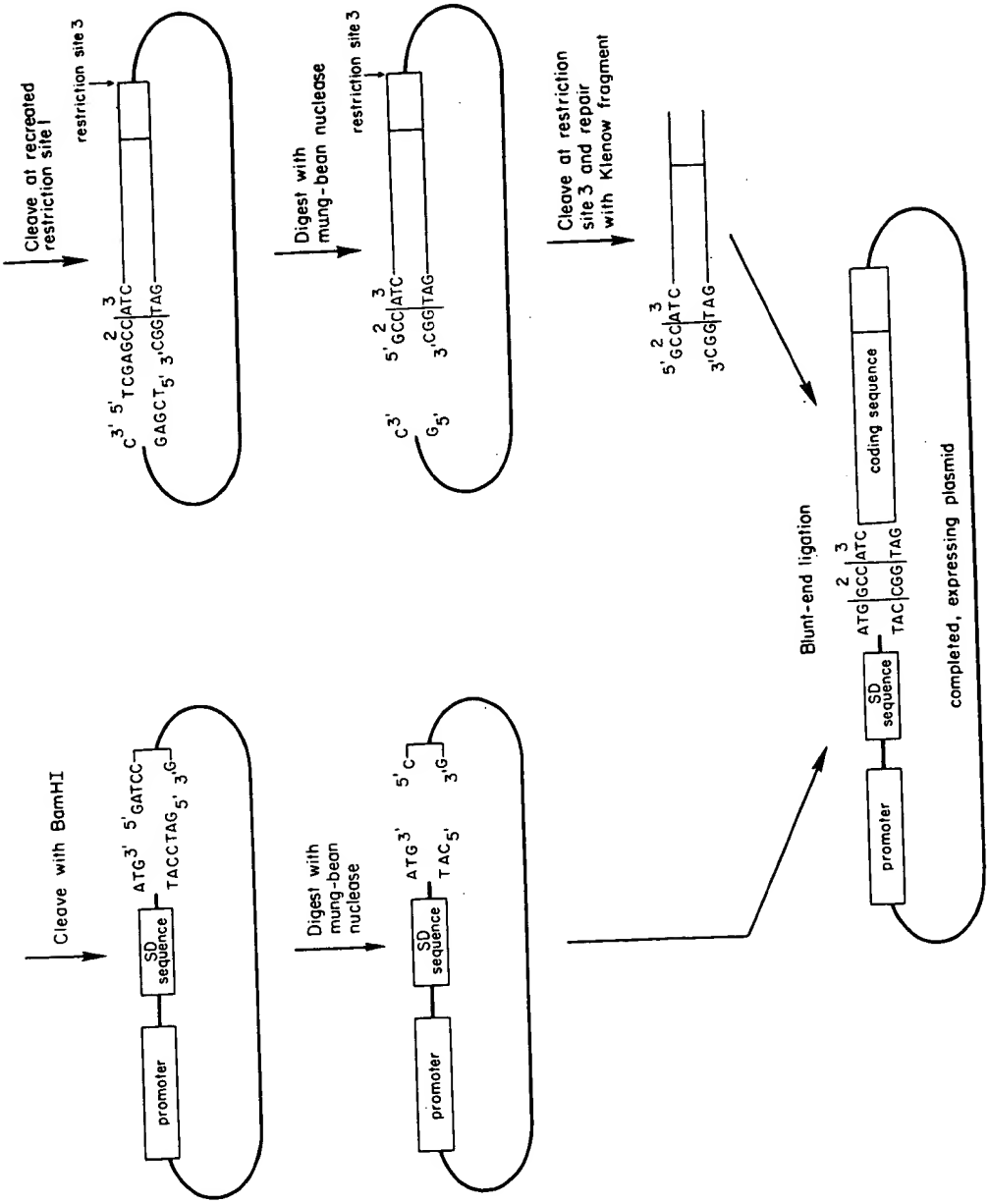


Figure 12.8
A method to create a blunt end immediately before a particular nucleotide in a segment of cloned DNA (see page 418 for details).

Vectors That Express Fused Eukaryotic Proteins

In some cases eukaryotic DNA sequences have been expressed as part of a fusion protein whose amino terminus is encoded by prokaryotic sequences and whose carboxyl terminus is encoded by eukaryotic sequences. Although usually less desirable than the unaltered eukaryotic protein, fused proteins have some useful properties:

1. Many of them are more stable in bacteria than the native eukaryotic protein (Itakura et al. 1977; Goeddel et al. 1979b; Davis et al. 1981).

2. The fusion protein may be secreted by the expressing bacteria if the eukaryotic DNA is attached to a bacterial sequence coding for a signal peptide that causes export of proteins to be translocated across membranes (Talmadge et al. 1980). However, this has been demonstrated in only one case.

3. In some cases, fusion proteins can be chemically treated to release the eukaryotic peptide in a biologically active form (Goeddel et al. 1979b; Shine et al. 1980).

4. Fusion proteins may be used as antigens (e.g., Kleid et al. 1981).

For eukaryotic sequences to be expressed, the translational reading frame must coincide with that of the prokaryotic gene to which they are fused. Some vectors are available that contain restriction cleavage sites in all three reading frames of the prokaryotic sequence (e.g., Charnay et al. 1978; Talmadge et al. 1980; Tacon et al. 1980; T. Silhavy, pers. comm.). Other vectors require the addition of synthetic linkers to allow the eukaryotic gene to be inserted in-frame (Shine et al. 1980). In either case, the DNA encoding the gene to be expressed must contain a restriction site that can be used to make the fusion; if a convenient site is not present, one must be inserted by exonuclease treatment followed by linker insertion.

pOP203-13

The plasmid pOP203-13 (Fuller 1981; see Fig. 12.9) is useful for constructing gene fusions whose products contain only a few amino acids encoded by a prokaryotic sequence (e.g., Fraser and Bruce 1978). Insertion of the coding sequence to be expressed into the *Eco*RI site of pOP203-13 allows a fusion protein to be expressed that contains the first 7 amino acids of β-galactosidase, a few amino acids encoded by *Eco*RI linker, and the amino acids of the eukaryotic polypeptide. The expression vector provides not only a *lac*uv5 promoter, which can be regulated by *lac* repressor in a *lacI*^{q} strain of *E. coli*, but also the ribosome-binding site of the *lacZ* gene. Because only a small number of amino acids of β-galactosidase are added to the protein to be expressed, fusion proteins synthesized using this vector may more nearly

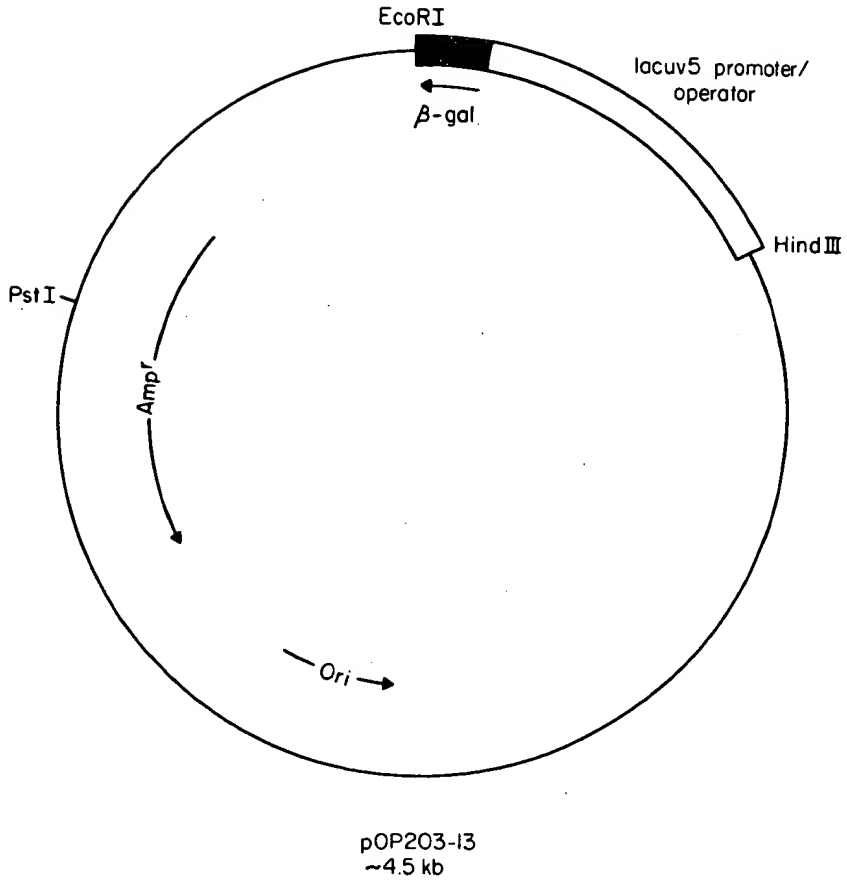


Figure 12.9

pOP203-13, a plasmid ~ 4.5 kb in length carrying the *lac*uv5 promoter and a single *Eco*RI site. Sequences inserted in-frame into this site will be translated into a fusion protein containing 7 amino acids of β-galactosidase and a few amino acids encoded by the *Eco*RI linker at their amino termini (Fuller 1981). This plasmid has been used to express a β-galactosidase-chicken-ovalbumin fusion protein (Fraser and Bruce 1978) and a fusion between β-galactosidase and human influenza virus hemagglutinin (Heiland and Gething 1981).

resemble the native protein than fusion proteins synthesized using the vectors described below. pOP203-13 has been used to express in *E. coli* a protein that consists of the N-terminal sequences of β-galactosidase fused to a large segment of influenza virus hemagglutinin (Heiland and Gething 1981).

pLC24

This plasmid (Figure 12.10) contains the strong p_L promoter of bacteriophage λ and directs the synthesis of fusion proteins consisting of 98 amino acids of MS2 polymerase and a polypeptide encoded by the foreign DNA sequence inserted at *Bam*HI or *Hind*III sites. Fusion proteins synthesized in this vector include MS2 polymerase-human fibroblast interferon (Derynck et al. 1980) and MS2-polymerase-foot-and-mouth-disease virus VP1 (Küpper et al. 1981). Expression of these fusion proteins may be conveniently controlled by regulating the activity of the p_L promoter in a lysogen with a λ*cIts*857 prophage.

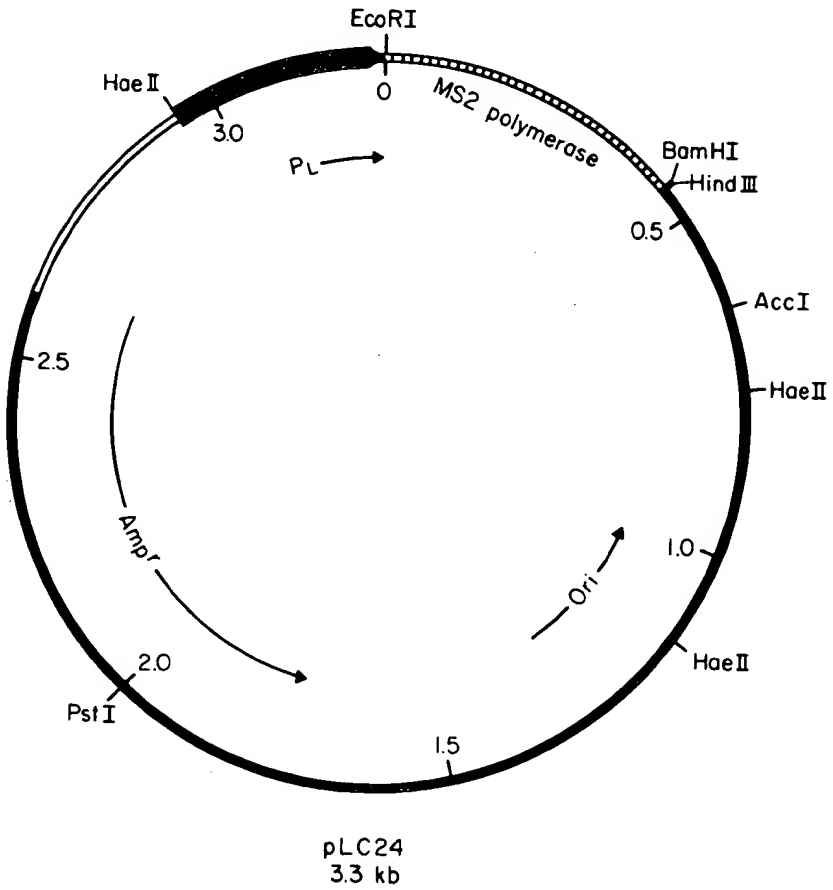


Figure 12.10

pLC24, a plasmid 3.3 kb in length that carries the bacteriophage λ p_L promoter and the ribosome-binding site of the bacteriophage MS2 polymerase gene. A fusion protein containing 98 amino acids of MS2 polymerase is expressed when a foreign gene is inserted in-frame into the *Bam*HI or *Hind*III sites. As in the case of other p_L promoter plasmids, inserted genes can be regulated by temperature in a λ*cIts*857 lysogen (Remaut et al. 1981). pLC24 has been used for expression of fusion proteins MS2 polymerase with human fibroblast interferon (Derynck et al. 1980) and with foot-and-mouth-disease virus VP1 (Küpper et al. 1981).

p*trpED*5-1

This plasmid (Figure 12.11) was designed to express a fusion protein composed partly of *trpD* and partly of a foreign polypeptide whose coding sequences have been inserted into the *Hin*dIII site of *trpD* (Hallewell and Emtage 1980). p*trpED*5-1 has been modified so that the foreign sequences can be inserted into each of the three reading frames (Tacon et al. 1980). With the *trp* attenuator present, very little transcription occurs in repressed conditions, so that this plasmid is useful for cloning genes whose products are toxic to *E. coli*. A fusion protein made from this vector contains about 15% of the sequence of the *trpD* polypeptide. The truncated *trpD* fragment is itself stable in *E. coli* cells, perhaps because it associates with the *trpE*-gene product (Hallewell and Emtage 1980). The fusion of the *trpD* polypeptide and a foreign gene product may also be stabilized by a similar association with *trpE*.

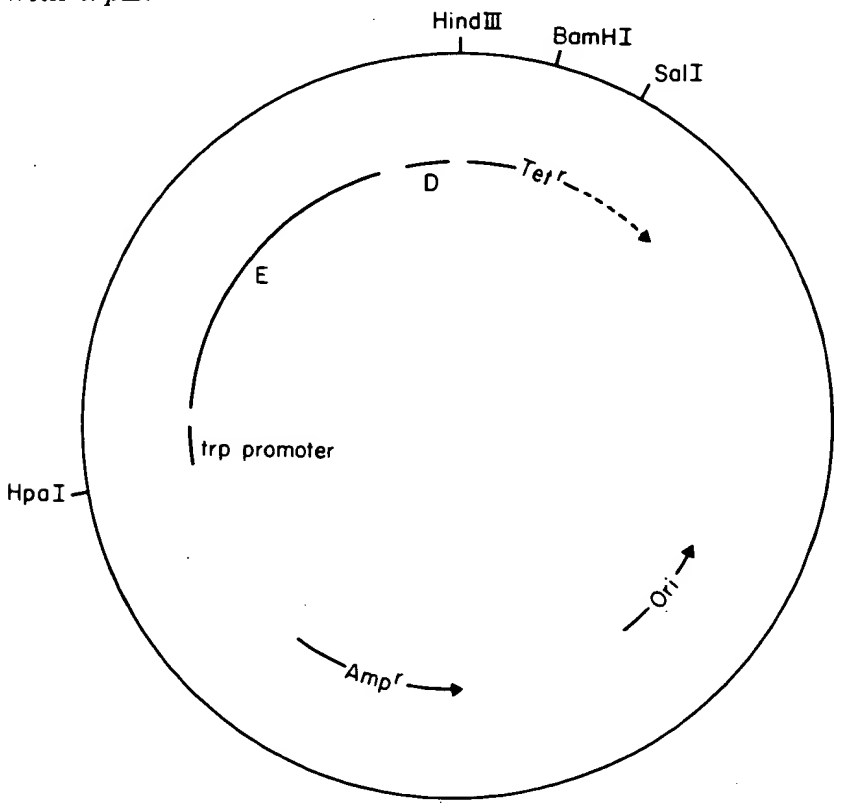


Figure 12.11

p*trpED*5-1, a plasmid ~6.7 kb in length carrying a *trp* promoter and designed to produce fusion proteins containing the aminoterminal 15% (75 amino acids) of the *trpD* protein. Induction of the *trp* operon with 3-indolylacetic acid results in at least a 50-fold increase in *trp*-gene products. Under conditions of repression, the *trp* proteins (and any fusion proteins derived from them) are expressed at a relatively low level (the *trp* attenuator is present in this plasmid). Genes are inserted at the *Hin*dIII site to allow expression of fusion proteins that may be relatively stable in *E. coli* (Hallewell and Emtage 1980). A different plasmid exists for each of the three reading frames (Tacon et al. 1980).

pNCV

Two other expression vectors have been designed to allow the synthesis of large fusion proteins that also may stabilize foreign gene products. One, plasmid pNCV (see Fig. 12.12), has been used to produce stable human influenza virus hemagglutinin (Davis et al. 1981) and foot-and-mouth-disease virus VP3 antigens (Kleid et al. 1981). The stability is a consequence of fusing the *trpLE* protein encoded by *trpΔLE* 1413 with the viral protein (Goeddel et al. 1980b). This vector was constructed by removing the termination codon of *trpLE* and replacing it with synthetic DNA encoding two *Eco*RI sites and a *Pst*I site. The sequence coding for a foreign protein may be inserted at one of these sites. This vector lacks the *trp* attenuator region, and the strong *trp* promoter is therefore always partially derepressed even in the presence of excess tryptophan.

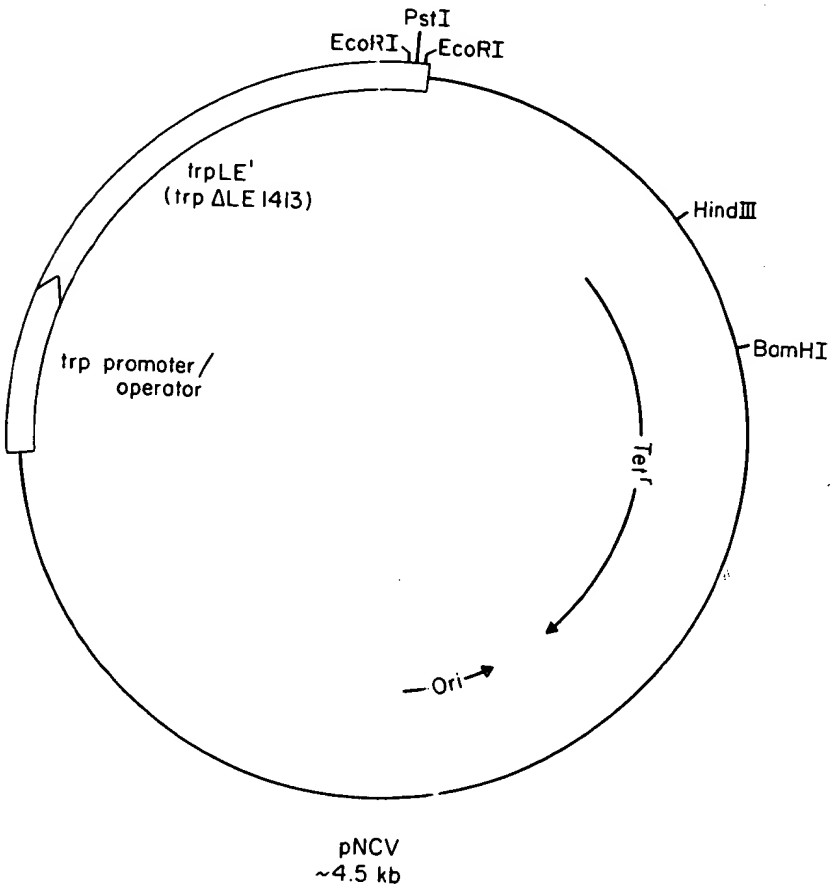


Figure 12.12

pNCV, a plasmid ~4.5 kb in length that allows synthesis of fusion proteins containing most of the *trpLE* fusion protein. The gene of interest may be inserted into the unique *Eco*RI or *Pst*I sites. The *trp* promoter lacks the attenuator region and is partially induced even in the presence of excess tryptophan. The vector has been used to produce stable fusion proteins of the *trpLE* protein with human leukocyte interferon (Goeddel et al. 1980b), with a human influenza virus hemagglutinin (Davis et al. 1981), and with foot-and-mouth-disease virus VP3 (Kleid et al. 1981).

pβ-*gal*13C

Another expression vector employing a large fusion protein for stabilizing foreign gene products is pβ-*gal*13C (see Fig. 12.13; Goeddel et al. 1979b). By cloning foreign DNA sequences into the single *Eco*RI site in *lacZ*, a fusion is made to the first 1005 amino acids of β-galactosidase (Goeddel et al. 1979b). A similar vector has been used to synthesize a hybrid β-galactosidase/β-endorphin protein (Shine et al. 1980).

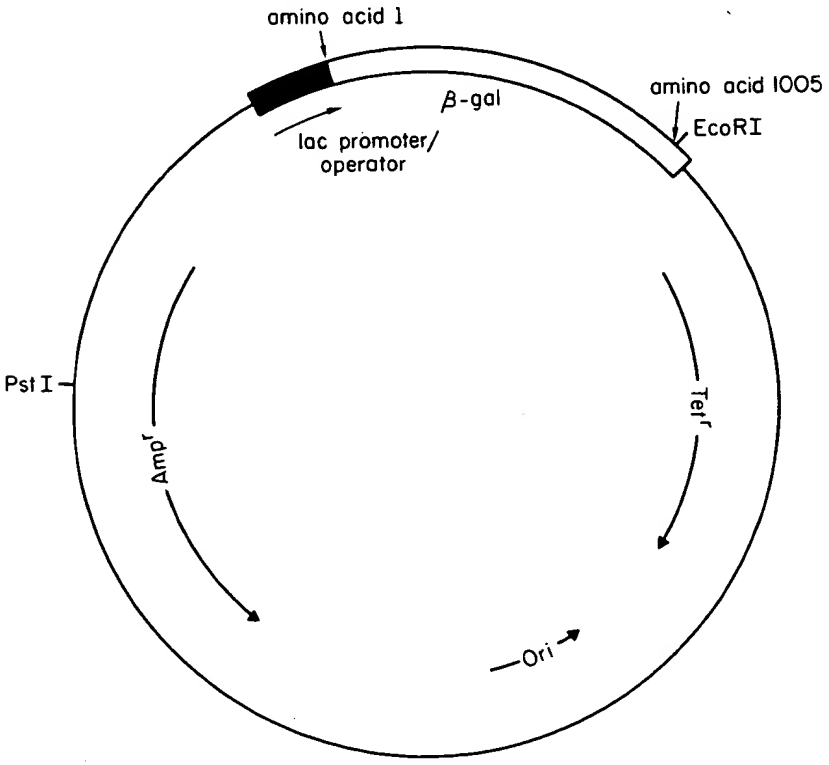


Figure 12.13

pβ-*gal*13C, a plasmid ~7 kb in length that contains a large segment of the *lacZ* gene, which encodes β-galactosidase. Insertion of DNA fragments in-frame into the unique *Eco*RI site may be more stable than the native eukaryotic protein in bacteria. Vectors designed for expression of genes as fusion proteins with 1005 amino acids of β-galactosidase at their amino terminus have been used to produce fusions with human insulin peptides (Goeddel et al. 1979b), with antigenic determinants of human influenza virus hemagglutinin (Davis et al. 1981), with somatostatin (Itakura et al. 1977), and with hepatitis B virus surface antigen (Charnay et al. 1980). Several β-galactosidase fusion proteins appear to be insoluble.

pKT287

The use of fusion proteins as a means to export foreign proteins from bacteria is still under investigation. A series of vectors has been constructed that allows insertion of DNA sequences in all three reading frames at a *Pst*I site in the *bla* gene of pBR322 (Fig. 12.14; Talmadge et al. 1980). Fusion proteins containing various amounts of the amino terminus of penicillinase (4 to 27 amino acids) may be made from these vectors. The first 23 amino acids encoded by *bla* comprise the β-lactamase signal sequence. A proinsulin sequence fused to this leader resulted in expression of a protein, 50% or more of which was processed and secreted to the periplasm (Talmadge et al. 1980). The signal sequence normally found on proinsulin also functions in *E. coli* to allow the secretion of proinsulin (Talmadge et al. 1981). However, some other eukaryotic proteins that have leader sequences (e.g., human pre-β-interferon) are not processed in bacteria (Taniguchi et al. 1980b).

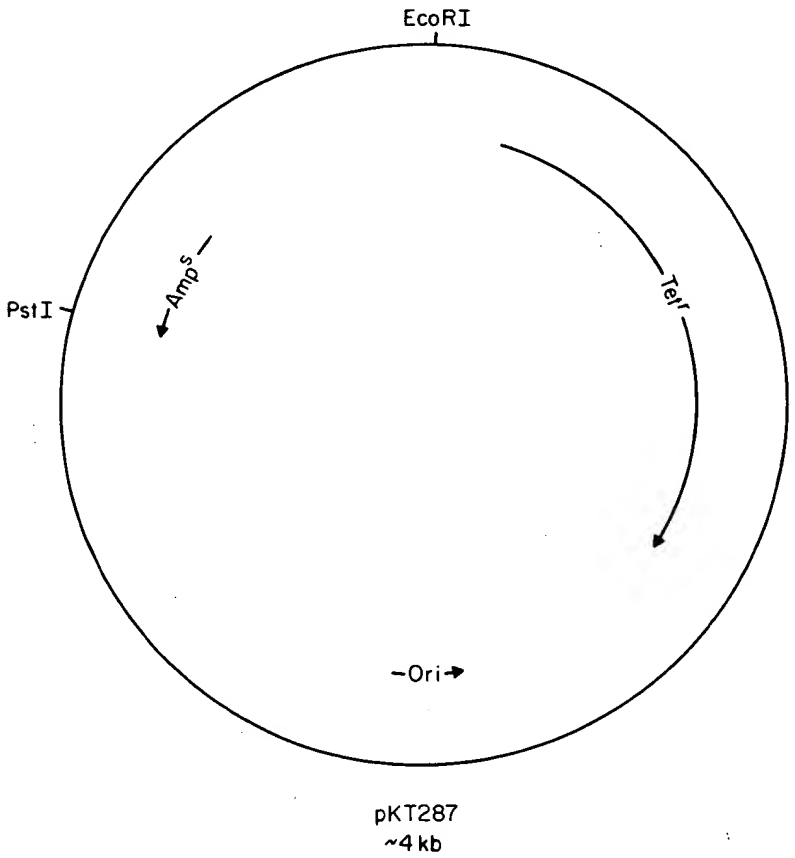


Figure 12.14

pKT287, a plasmid ~ 4 kb in length containing the promoter of the β-lactamase gene. If a gene is inserted in-frame at the *Pst*I site, a fusion protein is produced containing the leader sequence. Other plasmids in which the *Pst*I site is located at various positions within or downstream from the sequences encoding the signal peptide have also been constructed. These plasmids were made by cleaving pBR322 at the *Pst*I site, digesting the DNA with *Ba*l31, and inserting a *Pst*I linker (Talmadge et al. 1980).

pMH621

This vector also has the potential to cause the secretion of fusion proteins from bacteria (Fig. 12.15) (Hall and Silhavy 1981a,b; T. Silhavy, pers. comm.). When a foreign coding sequence is inserted at the *Bgl*II site, a fusion protein that has the *E. coli ompF* signal sequence is synthesized, as well as 12 amino acids of the mature *ompF* protein attached to the amino acids of the foreign polypeptide. Several polylinkers (see Chapter 1) have been inserted at the *Bgl*II site of pMH621 to allow insertion of foreign DNA using different restriction sites, to facilitate insertion in the correct reading frame, and to allow directional insertion. This vector is designed to allow the synthesis of large amounts of a fusion protein (the outer membrane protein encoded by *ompF* is present up to a level of about 100,000 copies per cell). In a bacterial strain with a cold-sensitive mutation in a positive regulatory gene (*ompR*), high levels of expression of the fusion protein occur only at high temperature (expression is minimal at 30°C). Because the hybrid proteins will contain the *ompF* signal sequence, they will probably be exported from the cytoplasm. However, it is not known whether the signal sequence will be sufficient for export. By analogy with the β-lactamase vector described above (pKT287), pMH621 may allow production of mature proteins whose leaders have been processed. This system may even export proteins to the cell surface.

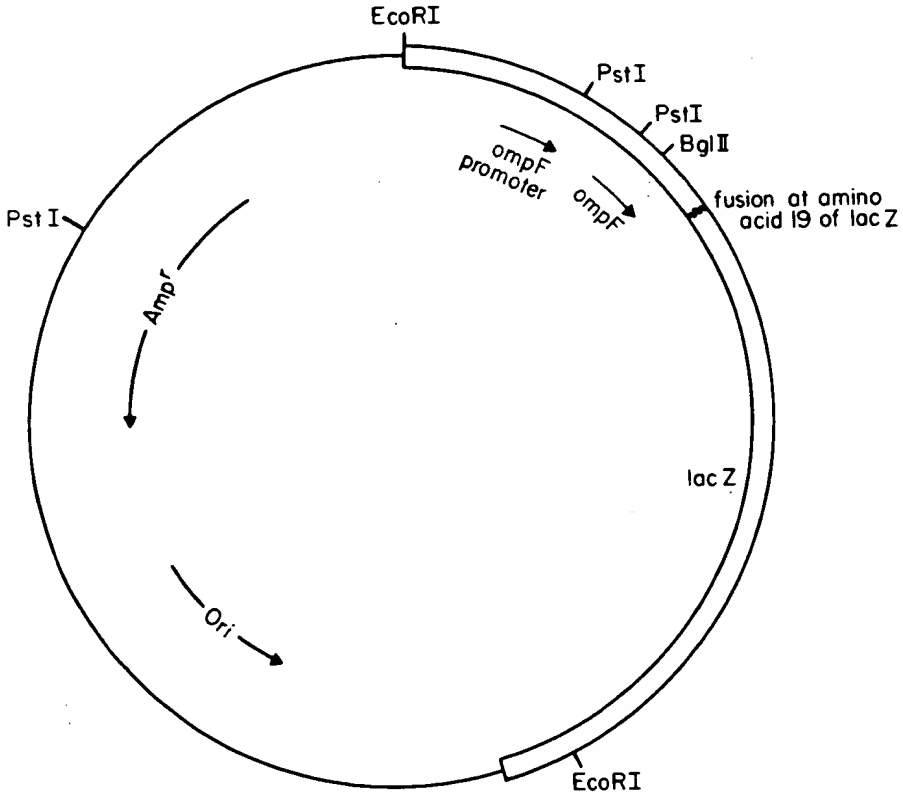


Figure 12.15

pMH621, a plasmid ~7.6 kb in length carrying the *ompF* promoter (*ompF* is a gene coding for a major outer membrane protein present at levels up to 100,000 copies per cell). Inserts in-frame into the *Bgl*II site produce a fusion protein containing the *ompF* signal peptide and 12 N-terminal amino acids of mature *ompF*. Such fusion proteins may possibly be exported to the outer membrane of *E. coli*. The promoter is under positive control and mutants have been isolated that allow expression from the *ompF* promoter to be regulated by temperature shift (T. Silhavy, pers. comm.; for background information, see Hall and Silhavy 1981a,b).

MAXIMIZING EXPRESSION OF CLONED GENES

The levels of expression of a foreign gene in *E. coli* are usually monitored by an appropriate functional assay (Struhl and Davis 1977; Chang et al. 1978), or by various immunological assays (Broome and Gilbert 1978; Heiland and Gething 1981). However, when attempting to maximize the expression of a gene product, such assays are cumbersome and can be expensive. To overcome this problem when an unfused eukaryotic gene product is to be expressed, a method in which the foreign DNA is fused to an appropriate fragment of the *lacZ* gene (Guarente et al. 1980b) is used. This method allows plasmids to be identified in which the cloned gene is efficiently transcribed and translated, even if the protein encoded by the cloned gene has no assayable activity. This technique employs a plasmid (pLG) containing *lacZ* DNA encoding an enzymatically active carboxyterminal fragment of β-galactosidase. No β-galactosidase is synthesized, however, because the promoter for the gene and the SD sequence are both absent from the plasmid. The gene to be expressed is fused in-frame to the 5′ end of the *lacZ* gene. The resulting fused gene encodes a hybrid protein that in most cases has β-galactosidase activity. However, this protein is not expressed because it still lacks a promoter and, if the upstream sequence is derived from a eukaryotic gene, an SD sequence. A portable promoter fragment can then be properly positioned in front of the fused gene as described above. Those promoter fragments that are optimally positioned should direct the synthesis of high levels of a fusion protein that has β-galactosidase activity. The plasmids with optimally placed promoter fragments can be recognized by transforming Lac^- bacteria and scoring for β-galactosidase activity on lactose indicator plates. These plasmids can then be used to reconstitute an unfused eukaryotic gene that is expressed at high levels (Guarente et al. 1980a,b).

INCREASED GENE DOSAGE

Once a gene has been expressed at high levels, a DNA fragment containing the entire assembly (consisting of the promoter, ribosome-binding site, and gene) may be transferred to a plasmid that can conditionally attain very high copy number in the absence of chloramphenicol amplification. The temperature-inducible "runaway replication" vector pKN402 (Uhlin et al. 1979), a small derivative of R1*drd*19, was designed for this purpose. Several derivatives of pKN402 have been constructed that provide additional cloning sites allowing one to score for insertion of foreign DNA (Bittner and Vapnek 1981). One derivative of pKN402, the vector pAS2, has proven useful for production of large amounts of protein from a cloned gene (Brent and Ptashne 1981; A. Poteete, unpubl.; A. Sancar, pers. comm.). A DNA fragment containing the *lac* or *tac* promoter and the expressed gene was cloned into pAS2. Cultures of a *lacI*^{q} strain carrying the plasmid were grown at 42°C for a few hours to increase plasmid copy number to more than 1000 copies per cell, and then isopropyl-β-D-thiogalactoside was added to inactivate the *lac* repressor. Very large increases in the production of the cloned gene product were obtained with this technique. A thermoinducible bacteriophage λ–ColE1 chimeric plasmid, pKC16 (Rao and Rogers 1978), may be used in a similar manner. This plasmid contains genes for ampicillin and kanamycin resistance, a region of bacteriophage λ DNA (from λ*N* through λ*P*, including the replication origin and a λ*cIts*857 gene encoding a temperature-sensitive repressor), and a ColE1 segment including the replication and immunity regions. When the temperature of the bacterial culture is raised to 42°C, λ transcription is turned on, the λ replication system is used, and the plasmid copy number increases to about 250 copies per cell (Rao and Rogers 1978). Large increases in the production of exonuclease III were obtained by cloning the *exo*III gene into pKC16 and then growing cells containing the recombinant plasmid at 32°C, followed by shifting the culture to 42°C (Rao and Rogers 1978).

SUMMARY

A number of expression vectors, some of which have been described here, have been designed to direct efficient transcription and translation of cloned genes. Several strong promoters have been employed, and because no single method has been used to compare their efficiencies, the choice among promoters is still somewhat arbitrary. The ability to regulate transcription is an important consideration, since the expression of gene products at high levels may be toxic to *E. coli* cells.

The signals in the DNA that direct efficient translation may be provided in several ways. If the foreign gene is to be expressed as a native, unfused protein, it may be (1) positioned so that its ATG is an optimal distance from the bacterial SD sequence by scoring for maximum expression (for review, see Guarente et al. 1980a), (2) placed downstream from an SD sequence so that its ATG is a predetermined distance from the SD sequence (e.g., Goeddel et al. 1979a), or (3) fused to an ATG present on the vector (A. Shatzman and M. Rosenberg, pers. comm.). Neither the effect of substituting nucleotides between the SD sequence and the ATG of the ribosome-binding site nor the result of changing the DNA sequence after the ATG has been examined in a systematic fashion. However, the spacing between the SD sequence and the ATG is known to affect the amount of protein synthesized (Backman and Ptashne 1978).

If the native protein is unstable in *E. coli*, expression of the gene product as part of a fusion protein may be desirable, especially if the native protein can be chemically cleaved from the purified fusion protein (e.g., Goeddel et al. 1979b) or if the protein is to be used to elicit the production of antibodies (e.g., Kleid et al. 1981). Alternatively, some proteins may be stabilized by synthesis in large amounts in *E. coli* (Shimatake and Rosenberg 1981).

Vectors that allow fusion of foreign genes to DNA encoding a signal sequence may be useful for exporting proteins out of the cytoplasm, especially if the signal peptide is cleaved during export of the protein. Export of the proteins may assist in subsequent purification and may serve to isolate them from cytoplasmic proteases. However, the factors that determine whether a given protein will be secreted when it is fused with a particular leader peptide have not been elucidated.

In most cases, the levels of expression of eukaryotic genes in *E. coli* are less than expected. In addition to factors that have been discussed, expression levels may be influenced by mRNA secondary structure and stability or by codon usage. In any case, conditions necessary for optimization of expression have not been generalized, so at present each new protein poses a different problem.